EXHIBIT D

Partially indexed items in Content Search in Office 365

05/10/2018   11 minutes to read   Contributors 👤 👤

In this article

File types not indexed for search

Messages and documents with partially indexed file types can be returned in search results

Partially indexed items included in the search results

Partially indexed items excluded from the search results

Indexing limits for messages in Content Search

More information about partially indexed items

See also

A Content Search that you run from the Office 365 Security & Compliance Center automatically includes partially indexed items in the estimated search results when you run a search. Partially indexed items are Exchange mailbox items and documents on SharePoint and OneDrive for Business sites that for some reason weren't completely indexed for search. In Exchange, a partially indexed item typically contains a file—of a file type that can't be indexed—that is attached to an email message. Here are some other reasons why items can't be indexed for search and are returned as partially indexed items when you run a search:

- The file type is unrecognized or unsupported for indexing.

- Messages have an attached file without a valid handler, such as image files; this is the most common cause of partially indexed email items.

- The file type is supported for indexing but an indexing error occurred for a specific file.

- Too many files attached to an email message.

- A file attached to an email message is too large.

- A file is encrypted with non-Microsoft technologies.

- A file is password-protected.

ⓘ **Note**

Most Office 365 organizations have less than 1% of content by volume and less than 12% by size that is partially indexed. The reason for the difference between volume and size is that larger files have a higher probability of containing content that can't be completely indexed.

For legal investigations, your organization may be required to review partially indexed items. You can also specify whether to include partially indexed items when you export search results to a local computer or when you prepare the results for analysis with Office 365 Advanced eDiscovery. For more information, see Investigating partially indexed items in Office 365 eDiscovery.

File types not indexed for search

Certain types of files, such as Bitmap or MP3 files, don't contain content that can be indexed. As a result, the search indexing servers in Exchange and SharePoint don't perform full-text indexing on these types of files. These types of files are considered to be unsupported file types. There are also file types for which full-text indexing has been disabled, either by default or by an administrator. Unsupported and disabled file types are labeled as unindexed items in Content Searches. As previously stated, partially indexed items can be included in the set of search results when you run a search, export the search results to a local computer, or prepare search results for Advanced eDiscovery.

For a list of supported and disabled file formats, see the following topics:

- **Exchange** - File formats indexed by Exchange Search

- **Exchange** - Get-SearchDocumentFormat

- **SharePoint** - Default crawled file name extensions and parsed file types in SharePoint

Messages and documents with partially indexed file types can be returned in search results

Not every email message with an partially indexed file attachment or every partially indexed SharePoint document is automatically returned as an partially indexed item. That's because other message or document properties, such as the **Subject** property in email messages and the **Title** or **Author** properties for documents are indexed and available to be searched. For example, a keyword search for "financial" will return items with an partially indexed file attachment if that keyword appears in the subject of an email message or in the file name or title of a document.

However, if the keyword appears only in the body of the file, the message or document would be returned as a partially indexed item.

Similarly, messages with partially indexed file attachments and documents of an partially indexed file type are included in search results when other message or document properties, which are indexed and searchable, meet the search criteria. Message properties that are indexed for search include sent and received dates, sender and recipient, the file name of an attachment, and text in the message body. Document properties indexed for search include created and modified dates. So even though a message attachment may be an partially indexed item, the message will be included in the regular search results if the value of other message or document properties matches the search criteria.

For a list of email and document properties that you can search for by using the Search feature in the Security & Compliance Center, see Keyword queries and search conditions for Content Search.

Partially indexed items included in the search results

Your organization might be required to identify and perform additional analysis on partially indexed items to determine what they are, what they contain, and whether they're relevant to a specific investigation. As previously explained, the partially indexed items in the content locations that are searched are automatically included with the estimated search results. You have the option to include these partially indexed items when you export search results or prepare the search results for Advanced eDiscovery. To include partially indexed items when you're exporting search results or preparing them for Advanced eDiscovery, select one of the options to include items that have an unrecognized format, are encrypted, or weren't indexed for other reasons.

Keep the following in mind about partially indexed items:

- When you run a content search, the total number and size of partially indexed items (returned by the search query) are displayed in search statistics in the details pane, as labeled as "unindexed items".

- When you export search results and include partially indexed items, partially indexed Exchange items are exported to a separate PST file for each mailbox in which they are located, or as individual messages if you select the option to download Exchange items as messages. partially indexed SharePoint items are exported to a folder named **Uncrawlable**.

- If the search that you're exporting results from was a search of specific content locations or all content locations in your organization, only the unindexed items from content locations that

contain items that match the search criteria will be exported. In other words, if no search results are found in a mailbox or site, then any unindexed items in that mailbox or site won't be exported. The reason for this is that exporting partially indexed items from lots of locations in the organization might increase the likelihood of export errors and increase the time it takes to export and download the search results.

To export partially indexed items from all content locations for a search, configure the search to return all items (by removing any keywords from the search query) and then export only partially indexed items when you export the search results (by clicking **Only items that have an unrecognized format, are encrypted, or weren't indexed for other reasons** under **Output options**).

- If you choose to include all mailbox items in the search results, or if a search query doesn't specify any keywords or only specifies a date range, partially indexed items might not be copied to the PST file that contains the partially indexed items. This is because all items, including any partially indexed items, will be automatically included in the regular search results.

- Partially indexed items aren't available to be previewed. You have to export the search results to view partially indexed items returned by the search.

Partially indexed items excluded from the search results

If an item is partially indexed but it doesn't meet the search query criteria, it won't be included as a partially indexed item in the search results. In other words, the item is excluded from the search results. For example, let's say you run a search and don't include any keywords or properties because you want to include all content. But you include a date range condition for the query. If a partially indexed item falls outside of that date range, it won't be included as a partially indexed item. Date ranges are an effective way to exclude partially indexed items from your search results.

Similarly, if you choose to include partially indexed items when you export the results of a search, partially indexed items that were excluded from the search results won't be exported.

One exception to this rule is when you create a query-based hold that's associated with an eDiscovery case. If you create a query-based hold, all partially indexed items are placed on hold. This includes partially indexed items that don't match the search query criteria and partially indexed items that might fall outside of a date range condition. For more information about

creating query-based holds, see Step 4 in eDiscovery cases in the Office 365 Security & Compliance Center.

Indexing limits for messages in Content Search

The following table describes the indexing limits that might result in an email message being returned as a partially indexed item in a Content Search in Office 365.

For a list of indexing limits for SharePoint documents, see Search limits for SharePoint Online.

Indexing limit	Maximum value	Description
Maximum attachment size (excluding Excel files)	150 MB	The maximum size of an email attachment that will parse for indexing. Any attachment that's larger than this limit won't be parsed for indexing, and the message with the attachment will be marked as partially indexed. **Note:** Parsing is the process where the indexing service extracts text from the attachment, removes unnecessary characters like punctuation and spaces, and then divides the text into words (in a process called tokenization), that are then stored in the index.
Maximum size of Excel files	4 MB	The maximum size of an Excel file located on a site or attached to an email message that will be parsed for indexing. Any Excel file that's larger than this limit won't be parsed, and the file or the email the message with the file attachment will be marked as unindexed.
Maximum number of attachments	250	The maximum number of files attached to an email message that will be parsed for indexing. If a message has more than 250 attachments, the first 250 attachments are parsed and indexed, and the message is marked as partially indexed because it had additional attachments that weren't parsed.
Maximum attachment depth	30	The maximum number of nested attachments that are parsed. For example, if an email message has another message attached to it and the attached message has an attached Word document, the Word document and the attached message will be indexed. This behavior will continue for up to 30 nested attachments.
	0	An image that's attached to an email message is skipped by the parser and isn't indexed.

Indexing limit	Maximum value	Description
Maximum number of attached images		
Maximum time spent parsing an item	30 seconds	A maximum of 30 seconds is spent parsing an item for indexing. If the parsing time exceeds 30 seconds, the item is marked as partially indexed.
Maximum parser output	2 million characters	The maximum amount of text output from the parser that's indexed. For example, if the parser extracted 8 million characters from a document, only the first 2 million characters are indexed.
Maximum annotation tokens	2 million	When an email message is indexed, each word is annotated with different processing instructions that specify how that word should be indexed. Each set of processing instructions is called an annotation token. To maintain the quality of service in Office 365, there is a limit of 2 million annotation tokens for an email message.
Maximum body size in index	67 million characters	The total number of characters in the body of an email message and all its attachments. When an email message is indexed, all text in the body of the message and in all attachments is concatenated into a single string. The maximum size of this string that is indexed is 67 million characters.
Maximum unique tokens in body	1 million	As previously explained, tokens are the result of extracting text from content, removing punctuation and spaces, and then dividing it into words (called tokens) that are stored in the index. For example, the phrase "cat, mouse, bird, dog, dog" contains 5 tokens. But only 4 of these are unique tokens. There is a limit of 1 million unique tokens per email message, which helps prevent the index from getting too large with random tokens.

More information about partially indexed items

- As previously stated, because message and document properties and their metadata are indexed, a keyword search might return results if that keyword appears in the indexed metadata. However, that same keyword search might not return the same item if the keyword

only appears in the content of an item with an unsupported file type. In this case, the item would be returned as a partially indexed item.

- If a partially indexed item is included in the search results because it met the search query criteria (and wasn't excluded) then it won't be included as a partially indexed item in the estimated search statistics. Also, it won't be included with partially indexed items when you export search results.

- Although a file type is supported for indexing and is indexed, there can be indexing or search errors that will cause a file to be returned as a partially indexed item. For example, searching a very large Excel file might be partially successful (because the first 4 MB are indexed), but then fails because the file size limit is exceeded. In this case, it's possible that the same file is returned with the search results and as a partially indexed item.

- Attached files encrypted with Microsoft technologies are indexed and can be searched. Files encrypted with non-Microsoft technologies are partially indexed.

- Email messages encrypted with S/MIME are partially indexed. This includes encrypted messages with or without file attachments.

- Messages protected using Information Rights Management (IRM) are indexed and will be included in the search results if they match the search query.

See also

Investigating partially indexed items in Office 365 eDiscovery